Frequency Decomposition Techniques for Increased Discriminative 3D Facial Information Capture

Clinton Fookes, Jamie Cook, Sridha Sridharan
Image & Video Research Laboratory
Queensland University of Technology
GPO Box 2434 Brisbane QLD 4001 Australia
{c.fookes,j.cook,s.sridharan}@qut.edu.au

Massimo Tistarelli
Computer Vision Laboratory
University of Sassari
Piazza Duomo 6 Alghero 07041 Italy
tista@uniss.it

Abstract

Eigen-based techniques and other monolithic approaches to face recognition have long been a cornerstone in the face recognition community due to the high dimensionality of face images. Eigen-face techniques provide minimal reconstruction error and limit high-frequency content while linear discriminant-based techniques (fisher-faces) allow the construction of subspaces which preserve discriminatory information. This paper presents a frequency decomposition approach for improved face recognition performance utilising three well-known techniques: Wavelets; Gabor / Log-Gabor; and the Discrete Cosine Transform. Experimentation illustrates that frequency domain partitioning prior to dimensionality reduction increases the information available for classification and greatly increases face recognition performance for both eigen-face and fisher-face approaches.

1. Introduction

The monolithic approach to face recognition has been well represented in the literature and is backed by considerable evidence of holistic processing in humans [18]. The major problem encountered when processing the appearance of the face in a holistic manner is the curse of dimensionality. Digital cameras are utilising increasing numbers of receptors in their sensor arrays, generating a greater number of pixels which greatly increases the complexity of a recognition task. This increase in complexity occurs as the dimensionality of a face representation increases, so too does the number of training observations that are required to provide numerical stability to algorithms utilising such a representation.

When each pixel of a face image is treated as a dimension in a feature space, the amount of training data required quickly becomes intractable. There have been many approaches to overcome this but by far the most popular are the subspace projection techniques. Broadly these can be grouped into two main categories,

- Supervised: training data is labeled and grouped into classes. This includes the Discriminant techniques such as Linear Discriminant Analysis (LDA) and Generalised Disriminant Analysis (GDA) which calculate transformations which best separate the classes present in the training data. This ensures that information lost during transformation is that which is least useful to class separability, but performance is limited by the quality of the training set [14].
- Unsupervised: in this case only exemplar images are presented for training, no knowledge of the class to which each image belongs is assumed. Transformations generated by this category will attempt to best characterise the entire input space, typically by maximising the energy or variance retained in the subspace. This category covers the popular Principal Component Analysis (PCA) algorithm.

Costen *et al*. [5] established the existence of a preferred band in the frequency domain for the human recognition of faces. This band is cited as being between 8 and 16 cycles per face, where the term cycles per face is defined as "*the number of sinusoidal repetitions of a given width that can be placed within the eye-level width of the face*". To give the reader a scale of reference, Figure 1 shows two example 2D faces as information is removed from this band. Each row shows an image of two easily distinguishable subjects, in the second column downsampling has occurred but the maximum frequency retained is 16 cycles/face and the specified frequency band is still intact. In the last two columns the upper cutoff is decreased with the maximum frequency dropping to 12 and 8 cycles/face respectively. The ability to clearly distinguish the two subjects is significantly diminished as the identified band is removed.

It has been demonstrated that subspace methods, in particular the PCA method, are particularly resilient to the il-

lustrated degradations. In fact when the introduced degradation affects only the high frequency information in face images [14] the performance of such systems is surprisingly robust.

Figure 1. Intensity 2D images, shown in (a), of two easily distinguishable subjects are downsampled such that the highest frequency remaining in the image is: (b) 16 (c) 12 (d) 8 cycles/face.

This non linear response to frequency distortion is an aspect of holistic subspace projection which has been identified by many researchers [11, 2] but has yet to be accurately quantified. Given that there exists a frequency band containing a significant portion of the discriminative face information in the frequency domain there are unanswered questions regarding the frequency response of these subspace methods. In particular: how precisely can this band be localised? What proportion of discriminative information is contained in this band? Do various subspace projection methods remove or distort the information in this band during transformation? If so how can this behaviour be modified or preferably obviated? Furthermore, the Small Sample Size problem is one that is present in all techniques which use high dimensional image data. Until database sizes increase exponentially to match image dimensionality, methods for allieviating this shall be required.

It is the intent of this paper that answers to some of the above questions and a solution to the Small Sample Size problem can be achieved for the purpose of 3D facial recognition through the use of frequency decomposition techniques to partition the information content in the 3D face. By training subspace projections for each partition separately, the amount of information encoded by the transformation is reduced. Thus, the same training data can be used to train a series of specialised subspaces which more accurately describe the underlying information. The benefits of frequency domain decomposition combined with local classifier training shall be shown to improve the performance of whole face processing.

The remainder of the article is as follows. Section 2 presents the effects of subspace projection. Section 3 presents three decomposition techniques for frequency-domain partitioning. Section 4 presents experimental results and analysis on the face recognition grand challenge (FRGC) dataset and the paper is concluded in Section 5.

2. Frequency Effects of Subspace Projection

The process of considering a $m \times n$ dimensional grid of image intensities (or range) as a single mnx1 vector discards the spatial relationships between pixels. The process of projecting face images from such a representation into a subspace will obviously introduce some very non-linear effects on the 2D spatial frequency domain information content of the signal. To illustrate this, the average frequency response of the PCA reconstruction of 3D face images over the FRGC database shall be constructed as a function of the number of retained eigenvectors.

The frequency response of a transformation, $I' = H(I)$ can be evaluated by averaging the ratio $\mathfrak{F}(H(I))/\mathfrak{F}(I)$, where $\mathfrak{F}(I)$ is the 2D Fourier transform of the image I. In this case both $I = f(x, y)$ and $\mathfrak{F}(I) = g(u, v)$ are 2 dimensional functions which are difficult to visualise. Thus $\mathfrak{F}(I)$ is redefined as $\mathfrak{F}(I) = g(r, \theta)$ and $F(I) = g(r) = \int_0^{2\pi} |\mathfrak{F}(I)| d\theta$.

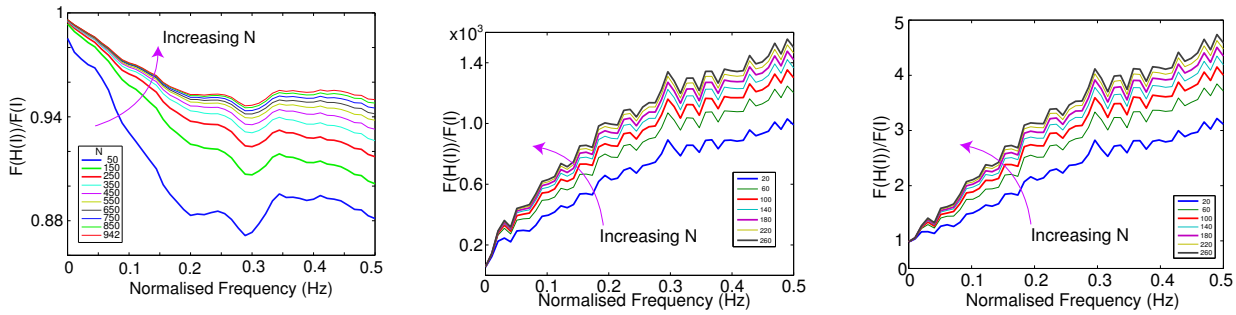
The system $H(I)$ being modeled is the reconstruction of a face image, I', after projection of the original image I into the subspace S_{PCA} which has N retained dimensions. Thus, $H(I|N) = V_N^T(I - \mu_I)V_N + \mu_I$, where V_N is the truncated transformation matrix of size $mn \times N$ and μ_I is the mean vector of the training set.

It can be seen from the averaged results shown in Figure 2(a) that as the number of retained eigenvectors is increased the response in high frequency regions is improved. However this improvement comes at the expense of extra dimensionality and even when using the full complement of eigenvectors the reconstruction still exhibits a low-pass characteristic.

This low-pass characteristic corroborates previous work which has shown that performance of Eigen-faces is very robust to downsampling and blurring of input images [2]. These are both operations which remove high spatial frequencies from an image and as shown these higher frequencies are also removed by PCA. Therefore it is unsurprising that removing them before calculating the transformation has little effect on the performance of such systems.

While being a requisite for real-world processing of data, this loss of information precludes classification algorithms from performing at full potential. To maximise discriminability it is desirable to reduce the dimensionality without significantly affecting the information content of the signal.

In contrast to the low-pass characteristic shown in the PCA transformation similar analysis of the LDA transformation space yields a very large high frequency bias. Results for the 3D responses are shown in Figure 2(b). The first point of interest in this representation is the introduction of energy into the frequency response. This is characterised by a frequency response with magnitude greater than unity; as can be seen from the displayed plots, the LDA re-



(a) (b) (c)

Figure 2. Approximated frequency response of (a) 3D Eigen-faces (PCA), (b) 2D Fisher-faces (LDA), and (c) PCA+LDA combination, as the number, N, of retained eigenvectors is increased respectively.

sponse lies in the range of 200-2000 times the original frequency magnitude. This is caused by the noisy nature of the LDA projection vectors. The LDA projection is optimised for discriminability rather than for reconstruction and hence it is entirely expected that the reconstructed images are not visually similar to the original images.

The reconstruction property can be improved by using an initial PCA dimensionality reduction stage before calculating the LDA transformation. This alternate form of the LDA transform changes the magnitudes of the response but not the fundamental shape as can be seen in Figure 2(c). By incorportating a PCA front end before LDA calculations the response now has a peak magnitude of less than 10 compared to a much higher peak in the previous case. A side by side comparison of the LDA and PCA+LDA frequency responses however, shows that they are equivalent up to a scaling factor.

It is the position of this paper that these presented non-linear effects on frequency response restrict the performance of subspace based methods. The use of frequency domain decomposition can improve performance by firstly restricting the non-linear effects to localised regions of the spectrum. Secondly, it can improve performance by reducing the inherent dimensionality and hence reducing the effect of the Small Sample Size problem.

3. Frequency Decomposition Techniques

The main objective of this paper is to investigate decompositional representations for 3D face data which can separate the information content prior to dimensionality reduction. This is motivated by biological systems; in human speech processing the Mel Frequency Cepstral Coefficients encode the information from a logarithmically spaced filterbank derived from the human auditory system [15]. In the visual domain, work done in the related field of Scale-Space [17] and wavelet theory [13] draws inspiration from the human visual system [8].

Gabor filters are commonly cited as sharing many properties with mammalian cortical cells [6] and appear to be a logical choice for the task of frequency partitioning. Wavelet analysis has also been used extensively to provide a compact representation which separates the frequency content of a signal or image. The Discrete Cosine Transform is often associated with compression tasks, however, its compactness and similarities to the Fourier Transform make it well suited to the current task. In the following section these methods are introduced in more detail and their suitability to the required task is investigated in detail.

To examine the efficacy of the proposed methodology, a standard nearest neighbor classication scheme is employed. Figure 3 shows a block diagram of the matching process given two input range images. After separation into frequency bands, each level is projected into a PCA subspace which has been trained with corresponding data. The subspace projections are compared using the mahalanobis cosine metric which was shown in [1] as having consistently high performance. The summation of distances across all levels is then used as the final matching score. The unweighted summation was chosen over more esoteric score fusion techniques as it has been shown that the increase in performance does not justify the extra complexity that such schemes require [12].

3.1. Discrete Wavelet Transform

Wavelet decomposition is the projection onto an orthonormal set of basis vectors which are generated by dilation and translation of a single "mother wavelet". The simplest such mother wavelet is the Haar wavelet,

$$\psi(t) = \begin{cases} 1 & : \quad 0 \leq t < 1/2 \\ -1 & : \quad 1/2 \leq t < 1 \\ 0 & : \quad \text{else} \end{cases}, \tag{1}$$

which generates the basis set,

$$\left\{ \psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - 2^j n}{2^j} \right) \right\}_{(j,n) \in Z^2}. \tag{2}$$

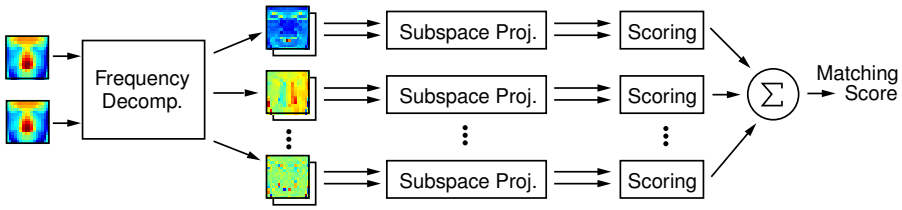


Figure 3. Block diagram of the matching process utilising frequency decomposition.

The Discrete Wavelet Transform (DWT) is a projection of a signal onto a finite number of such basis functions. It is typically implemented using two complementary filters derived from the "mother wavelet" which have low-pass and high-pass responses respectively. At each application of the filters the signal is split into a low and high frequency component which are then down-sampled by a factor of two to give consistent memory requirements regardless of decomposition level; this allows the same filters to be reused at each level as they have also effectively been up-scaled by a factor of 2. The multi-level representation which can be achieved using the DWT can be coerced into a filter bank form by grouping together the detail coefficients at each level. In the following experimentation 5 levels of decomposition are employed which yields the equivalent of 6 filter banks (5 detail vectors and the remaining approximation vector).

3.2. Gabor Filter Bank

Gabor filters share many properties with mammalian cortical simple cells such as spatial localisation, orientation selectivity and spatial frequency characterisation [6]. They have been successfully used in a variety of image processing applications ranging from texture analysis and expression characterisation to face recognition [4]. Perhaps most importantly they achieve the theoretical lower bound of joint uncertainty in the conjoint domains of visual space and frequency [7] allowing the construction of compact spatio-temporal filters. They can also be considered within the wavelet framework as the complete bank of Gabor filters can be constructed by dilations, rotations and translations of a single function. This functional form of the Gabor family of filters has two main components, the complex sinusoidal carrier, $s(x)$, and the gaussian envelope, $w_r(x)$. These are combined as,

$$\begin{aligned} \psi(x,y) &= s(x,y).w_r(x,y) \\ &= e^{2\pi j[u_0 x + v_0 y]}.Ke^{-\pi[x_r^2 a^2 + y_r^2 b^2]}, \end{aligned} \quad (3)$$

where u_0 and v_0 define the spatial frequency of the complex valued plane wave, a and b define the sharpness of the gaussian envelope. Also,

$$x_r = x\cos(\theta) + y\sin(\theta) \quad (4)$$

$$y_r = -x\sin(\theta) + y\cos(\theta), \quad (5)$$

where θ is the rotation of the gaussian about the origin, by convention this is normally chosen to be the same as the orientation of the filter in the frequency domain. In practice a compensation term, $e^{-\sigma^2/2}$, is commonly included to remove the DC component of the filters.

In [10] Field proposes an alternate method to perform both the DC compensation and to overcome the bandwidth limitation of a traditional Gabor filter. The Log-Gabor filter has a response that is gaussian when viewed on a logarithmic frequency scale instead of a linear one. This allows more information to be captured in the high frequency areas and also has desirable high pass characteristics. Field defines the frequency response of a Log-Gabor filter as,

$$\Phi(\omega) = \exp -\frac{\ln(\omega/\omega_0)}{2\ln(\sigma/\omega_0)}, \quad (6)$$

where ω_0 is the center frequency of the sinusoid and σ is a scaling factor of the bandwidth.

3.3. Discrete Cosine Transform

The Discrete Cosine Transform (DCT) is closely related to the Discrete Fourier Transform. Both the DCT and the Fast Fourier Transform (FFT) are separable, linear transforms, however, except for the special case where the signal under consideration is even, they are distinct. Instead of representing a signal, $f(x)$ by the weighted summation of complex sinusoids, $e^{j\omega x}$, the DCT instead uses a set of cosines, $\cos(\omega x)$, as basis functions. The 1D-DCT is defined as,

$$X_k = \sqrt{\frac{2}{\pi}} \left\{ \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}(n+\frac{1}{2})k\right] \right\}, \quad (7)$$

and due to the DCT's linear separable nature the 2D transform can be calculated by cascading two 1D-DCTs in the horizontal and vertical directions. The DCT has the effect of concentrating the energy in a signal in a relatively small number of coefficients and has been widely used in image compression schemes including the popular JPEG image format.

In order to construct filter-banks from a DCT representation a fundamental difference between the DCT and the other methods must first be addressed. This is namely that while the DWT and the Gabor filter response defines the

output of a particular filter at a given pixel the DCT instead defines the average response to the DCT basis vectors across the area in which it is applied. Thus the DCT cannot be applied across the entire image in the same manner as the previous methods. Instead, the common overlapping block implementation, as is used in the Part-Face method of Lucey [12] and Sanderson [16], is utilised. The coefficients from each block are divided into bands based upon the radial frequency of the corresponding basis vector. As with the DWT, this places a variable number of co-efficients into each level, with high-frequency bands containing more co-efficients than low.

4. Experimental Results

The three proposed methods are now evaluated for their suitability to the task of frequency domain partitioning. A standard nearest neighbour classification scheme is employed along with the Mahalanobis Cosine distance measure. Figure 3 shows a block diagram of the matching process given two input range images. Note that the following will use each of the proposed methods at the decomposition stage. The unweighted summation of distances across all levels is used as a matching score for preliminary combination analysis. The choice of an optimal combination strategy is not covered in these experiments, however, can further improve results when investigated. Evaluation of the three methods is conducted using 3D data from the Face Recognition Grand Challenge (FRGC) version 2.0 data set.

4.1. Discrete Wavelet Transform

The DWT is a dyadic multi-level transformation and as such provides a rigid partitioning of the frequency domain. At each level the frequency domain is approximately split in two by the high-pass and low-pass filters. The term approximately is used because the exact properties of the high and low pass filters depend upon the basis wavelet used.

The center frequencies of the DWT can be estimated by assuming that each iteration cuts the frequency domain into two equal sized regions. Different wavelets will of course have slightly different low-pass, high-pass characteristics but this serves well for the purpose of approximation. In the following all frequency measurements are given on the normalised frequency scale where $f_{sampling} = 1Hz$. The highest frequency which can be represented in this scale thus has a normalised value of 0.5Hz according to Nyquist sampling theory.

Assume that each level approximately halves the frequency domain and that the upper half is used to generate the detail coefficients. Given this, the detail coefficients encompass information from the range [0.25, 0.5] Hz and the center frequency of this would then be 0.375 Hz. The second level would encode the frequency range [0.125, 0.25] and have a center frequency of 0.1875 Hz, subsequent levels would have center frequencies which decrease in a similarly dyadic manner. The estimated center frequencies for the bands used in experimentation are listed in tabular form along with the resulting Equal Error Rates (EER) in Table 1.

The first point to notice from Table 1 is the negative trend in accuracy as the order of the wavelet is increased. This can be explained through examination of the mother wavelets for these wavelet families; in Figure 4 the wavelet functions for the lower orders of the Daubechies and Coiflet families are shown. One of the principal benefits of the wavelet transform is its ability to capture transient phenomena, as the wavelet order increases so too does the sharpness of the filter and the accuracy with which such phenomena can be localised. Compare then the sharpness of these filters against the smoothly changing nature of the human face and it becomes obvious that it is more appropriate to use the smoother lower order wavelets. While not shown here similar trends were observed in the Symlet families of wavelets and these results reflect those found in [3].

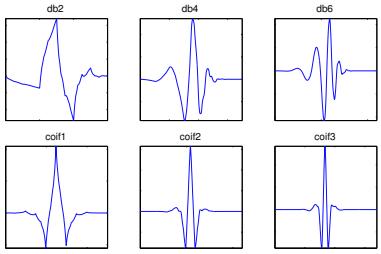


Figure 4. Ψ functions of Daubechies and Coiflet wavelets.

As outlined in the table, it was the Haar wavelet that gave the best performance across all frequency bands and the results from this wavelet is used to represent the DWT in subsequent comparisons with other frequency partitioning schemes. Another trend which is observed in the results is that very low and very high frequency bands do not generally perform as well as those in the middle. As the wavelet order is increased this is less apparent but in lower orders a clear dip can be seen for bands which have a center frequency of between 3 and 12 cycles/face. This lends some credence to the existence of a preferential band of discriminative information.

4.2. Gabor Filters

The Gabor and Log-Gabor filters are the closest in nature to the manner in which human's process visual stimulus [6] and allow the greatest flexibility in selecting partitions. The task of constructing an optimal filter bank for a recognition task is complicated and must consider a variety of factors. Where should center frequencies be located? How much overlap should there be between filters? What is an opti-

Center Frequency		Daubechies				Coiflet		
(cycles/pixel)	(cycles/face)	db1	db2	db4	db6	coif1	coif2	coif3
0.1875	24	13.56%	22.08%	27.06%	27.57%	21.53%	26.51%	27.41%
0.0938	12	9.86%	12.32%	15.54%	14.95%	12.71%	14.96%	15.11%
0.0469	6	7.77%	8.57%	9.68%	11.25%	10.11%	10.78%	11.22%
0.0234	3	8.57%	8.79%	9.20%	10.10%	9.59%	10.18%	8.99%
Approximation	1	13.39%	12.40%	12.61%	12.97%	12.94%	13.41%	13.39%
Combination	-	5.75%	6.09%	6.85%	7.91%	7.27%	7.38%	7.72%
		Bi-orthogonal						
		bior 1.3	bior 1.5	bior 2.2	bior 2.6	bior 3.1	bior 3.5	bior 3.9
0.1875	24	14.24%	14.69%	26.92%	27.26%	31.82%	31.15%	30.96%
0.0938	12	10.07%	10.35%	16.24%	15.30%	24.01%	19.80%	19.07%
0.0469	6	7.82%	7.86%	12.02%	9.72%	17.92%	12.42%	12.07%
0.0234	3	8.62%	8.67%	9.97%	8.14%	14.99%	10.35%	10.19%
Approximation	1	13.78%	13.89%	13.77%	12.24%	19.44%	13.47%	13.04%
Combination	-	5.94%	6.08%	7.93%	6.49%	14.94%	10.22%	8.63%

Table 1. Performance, expressed as EER (%), of Wavelet coefficient based representation on FRGC Experiment 3 3D dataset.

mal filter bandwidth? Interested readers requiring further information on optimal filter-bank construction are referred to [4].

In this experiment the center frequencies are fixed by the use of the less flexible DWT in the previous section. One aspect of filter-bank construction which is of central importance to the comparison between Gabor and Log-Gabor filters, however, is that of filter-bank overlap. Should filters have minimal overlap in order to increase the independence of extracted features? Or should they blanket the entire domain in order to ensure full coverage? Most work completed in the face recognition domain has used Gabor filters along with feature fusion to combine the various scales and orientations. Such an approach is benefited by having independent filters which reduce the redundant information and hence the required size of the filter bank.

In the proposed framework, classifier rather than feature fusion is used to combine the various scales, thus making filter independence less crucial. A stated benefit of the Log-Gabor family of filters is that by using them it is possible to construct filters of arbitrary bandwidth. When evaluating their suitability to the task of face recognition it is therefore required that consideration is given to the benefits of this characteristic. As previously stated the number of bands and their center frequencies are fixed, therefore only one variable parameter remains, bandwidth. By fixing these parameters, it is now possible to perform a systematic comparison of the filter families with particular respect to their differentiating characteristic, that of Log-Gabor filters to create arbitrary bandwidth filters.

Standard practise in creating filter banks is to use the point at which bandwidth is measured as the boundary between each filter in the bank. The fixation of center frequencies in a dyadic scale also fixes the measured bandwidth of each filter to be also dyadic, thus $\Delta F = 1$ octave. To keep with convention while enabling testing across a range of overlaps, the bandwidth shall be controlled implicitly via the parameter, α, which is defined as the percentage of peak magnitude at which bandwidth is measured. In this definition the -3dB bandwidth would be obtained by using $\alpha = 10^{-3/10} = 0.5012$.

The bandwidth of each filter in a bank of Log-Gabor filters is kept constant by defining the scaling factor $s = \sigma/\omega_0$, where ω_0 is the center frequency of the current filter. The bandwidth, $\Delta\omega$, of the filters, when measured in octaves, is directly related to the scaling factor by,

$$s = \exp\left(\pm\sqrt{\frac{-\ln^2(2^{\Delta\omega})}{8\ln\alpha}}\right). \tag{8}$$

For a Gabor family of filters, a similar constraint can be applied to limit the bandwidth in the radial direction,

$$\Delta\omega_\alpha = \log_2\left(\frac{\omega_0 + aC}{\omega_0 - aC}\right) \tag{9}$$

$$a = \frac{3\omega_0}{5}\sqrt{\frac{-\pi}{\ln\alpha}}. \tag{10}$$

This gives a unified mathematical representation which allows both families of filters to be controlled via one parameter. Testing will now show how the relative spread of the filters affects the performance of face recognition carried out using the output of the filter banks.

In the following the α parameter is considered on the range from [0.3,0.8]. Recognition rates across the database are calculated for each band of the filter bank individually and are displayed graphically in Figure 5. Results are shown for each band individually and also for the combined results obtained by taking the unweighted summation of each band.

The first point to note is that the use of additional overlap between bands appears to have a positive effect upon performance for most trialled center frequencies. This occurs for both Gabor and Log-Gabor filters, and is an expected result; as each filter bank is wider, more information is available to the classifiers leading to better performance. This effect is most noticeable in the lower frequency bands where the relative increase in bandwidth was higher. In the high frequency band ($f_0 = .1875$Hz), however, this trend was reversed and increasing overlap decreased performance. This

can be attributed to the inclusion of high frequency noise caused by the increased footprint of the filter.

Interestingly, there did not appear to be any significant change in the performance of the combination classifier, that is the classifier formed by summing the output of the five bands into a single score. While significant reductions were observed in all the individual classifiers their combination did not yield any additional benefits. This is caused by poaching, each classifier has additional information with which to make a verification decision, however that additional information is already represented in neighbouring classifiers.

When using the Log-Gabor filters the combination results seem to be fairly independent of the parameter α. The Gabor family exhibits slightly more variation with respect to α, in particular, the results deteriorated with increases in α. These results demonstrate a small but consistent improvement using Log-Gabor filters over Gabor filters, the best EER drops from 5.30%→4.86% for the FRGC 3D data. They also illustrate a very important behaviour of the Log-Gabor filters, their consistent performance across a range of the parameter α. The choice of filter parameters is a complicated process and the performance effects of poor choices can be quite dramatic. The fact that Log-Gabor filters exhibit robustness to a large range of choices in filter bandwidth highly recommend them as a method of feature extraction.

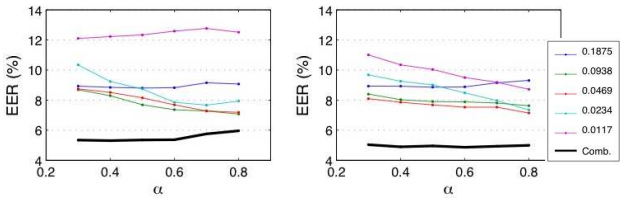


Figure 5. Effects of parameter α on recognition rates using both Gabor filters (left) and Log-Gabor filters (right). Results are grouped by center frequency of filter band and presented on 3D data from FRGC Experiment 3.

4.3. Discrete Cosine Transform

After selecting frequency bands from the DWT, and Gabor representations, equivalence with the DCT is established using the method outlined in Section 3.3. Each band of the Gabor filter defines a lower and upper cutoff from the α-magnitude profile which is fixed by the choices made for the DWT representation. By selecting coefficients from only the DCT basis functions which have peak frequency response within these bounds it is possible to establish equivalent representations using all three methods.

The block based nature of the DCT makes exact comparisons with the DWT and Gabor filters difficult. To make the comparison as equal as possible a range of block sizes were considered to provide differing levels of spatial locality to the extracted features. Also considered was the question of block overlap. By overlapping the blocks used for extraction it is less likely that discriminative features will be excluded from the representation but at the cost of introducing redundancy. To examine this both a non-overlapping approach and a 50% overlapping block approach is trialled.

This experimentation was conducted again using the FRGC Experiment 3 3D data and results are displayed in Table 2. The use of overlapping blocks appears to have a considerable performance advantage. This was especially noticeable with the 64 pixel block size, mainly due to the increased number of coefficients that are obtained in this manner. Further increasing the overlap to 75% was found to yield no further improvements to the reported recognition rates.

There seemed to be little to differentiate a block size of 32 from a block size of 64, however, when using a block size of 8 pixels, lower frequency information can not be encoded and this appears to have a serious impact on recognition accuracy. It is also notable that the DCT encoding of high frequency information is less suitable for recognition than was the Gabor representations examined in the previous section. Unlike experiments in the previous section where all bands contributed to the combined performance, the DCT high-frequency band degrades combined performance and is hence omitted from the combination results presented.

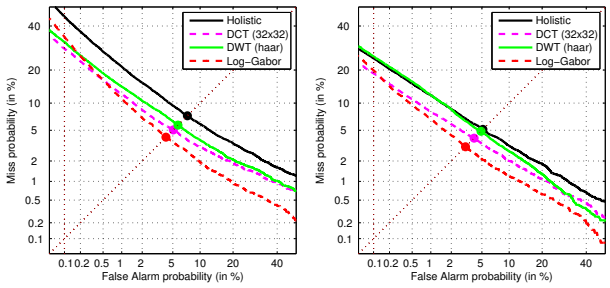
Center Frequency	Non-overlapping			Overlapping		
	16×16	32×32	64×64	16×16	32×32	64×64
0.1875	23.57%	25.77%	38.84%	23.17%	25.02%	25.78%
0.0938	10.95%	12.13%	21.52%	10.83%	11.68%	12.03%
0.0469	6.41%	7.56%	12.70%	6.55%	6.47%	7.03%
0.0234	-	8.92%	11.58%	-	6.70%	7.01%
0.0167	-	-	-	-	-	10.83%
Comb.	6.69%	5.86%	10.30%	6.71%	5.05%	5.14%

Table 2. Performance, expressed as EER (%), of DCT coefficient based representation on FRGC 3D face dataset.

4.4. Comparison

Previously it was posited that a frequency decomposition approach can improve the performance of subspace-based dimensionality by localising non-linear frequency distortions. To verify this the Detection Error Tradeoff (DET) curves of the top performing parameter set from each of the presented methods are now compared against a baseline system. The baseline system is comprised of a popular holistic method for face recognition, the Independent Component Analysis (ICA) algorithm, and the architecture II of [9] was selected. Results from the FRGC 3D data are shown in Figure 6 for both PCA- and LDA-based experimentation.

Immediately it is obvious that the frequency decomposition methods presented here dramatically improve upon the performance of the classic eigen-faces and fisher-faces ap-



(a) (b)

Figure 6. DET Comparison of (a) PCA and (b) LDA based techniques on the 3D FRGC database.

proaches. Large improvements can be observed with the Log-Gabor filters consistently providing the best performance. The relative improvement in EER for the 3D data utilising the PCA-based projection is 34% and for the LDA-based projection it is 39%.

These improvements are quite significant and illustrate the benefits of using specialised subspaces over a transformation which attempts to encode the entire frequency domain in a single projection. In all tested scenarios, the methods which trained subspaces on multiple bands of the spectrum all outperformed the baseline system. In addition to overall improvement in perfomance the Log-Gabor family of filters also offers a very important advantage over the other methods trialled, namely the superior flexibility in adjusting parameters.

5. Conclusion

This paper has presented a frequency decomposition approach for improved face recognition performance utilising three well-known techniques: Wavelets; Gabor and Log-Gabor; and the Discrete Cosine Transform. Experimentation has shown that frequency domain partitioning prior to dimensionality reduction increases the information available for classification and greatly increases face recognition performance. The three techniques were tested using data from the FRGC Experiment 3 dataset. Experimentation using unweighted classifier fusion has shown that partitioned data is capable of improving the performance of both traditional PCA and LDA subspace projection techniques. Relative improvements in equal-error rates for 3D data for the PCA-based projection and the LDA-based projection is 34% and 39% respectively. Furthermore, the performance of the Log-Gabor family of filters were shown to provide the superior advantage over other methods due to their flexibility in adjusting parameters.

References

[1] M. Bolme, R. Beveridge, and B. Draper. The csu face identication evaluation system: Its purpose, features and structure. In *Proceedings of the International Conference on Vision Systems*, pages 304–311, 2003. 3

[2] K. Chang, K. Bowyer, , and P. Flynn. Face recognition using 2d and 3d facial data. In *ACM Workshop on Multimodal User Authentication*, pages 25–32, 2003. 2

[3] J.-T. Chien and C.-C. Wu. Discriminant waveletfaces and nearest feature classifiers for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12):1644–1649, 2002. 5

[4] J. Cook, V. Chandran, and C. Fookes. 3d face recognition using log-gabor templates. In *British Machine Vision Conference*, Edinborugh, Scotland, 2006. 4, 6

[5] N. Costen, D. Parker, and I. Craw. Effects of high-pass and low-pass spatial filtering on face identification. *Percept Psychophys*, 58(4):602–12, 1996. 1

[6] J. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20:847–856, 1980. 3, 4, 5

[7] J. Daugman. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Trans. Acoust., Speech, Signal Process*, 36(7):1169–1179, 1988. 4

[8] G. DeAngelis, I. Ohzawa, and R. Freeman. Receptive-eld dynamics in the central visual pathways. *Trends Neuroscience*, 18(10):451–458, 1995. 3

[9] B. Draper, K. Baek, M. Bartlett, and J. Beveridge. Recognizing faces with pca and ica. *Computer Vision and Image Understanding*, 91(1-2):115–137, 2003. 7

[10] D. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of Optical Society of America*, 4(12):2379–2394, 1987. 4

[11] A. Lemieux and M. Parizeau. Experiments on eigenfaces robustness. In *Proceedings of the IEEE International Conference on Pattern Recognition*, volume 1, pages 421–424, Washington, DC, 2002. 2

[12] S. Lucey and T. Chen. Face recognition through mismatch driven representations of the face. In *Proceedings of the IEEE International Workshop on Visual Surveillance Performance Evaluation of Tracking and Surveillance*, pages 193–199, 2005. 3, 5

[13] S. Mallat. *A wavelet Tour of Signal Processing*. Academic, New York, 1998. 3

[14] A. Martinez and A. Kak. Pca versus lda. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):228–233, 2001. 1, 2

[15] L. Rabiner and B. Juang. *Fundamentals of Speech Recognition*. Prentice-Hall, Upper Saddle River, NJ, 1993. 3

[16] C. Sanderson and K. K. Paliwal. Fast features for face authentication under illumination direction changes. *Pattern Recognition Letters*, 24(14):2409–2419, 2003. 5

[17] J. Sporring, L. Florack, M. Nielsen, and P. Johansen. *Gaussian Scale-Space Theory*. Kluwer, Norwell, MA, 1997. 3

[18] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003. 1